5Ф

УДК 004.272

Л.П. Фельдман, Т.В. Михайлова
Донецкий национальный технический университет, г. Донецк, Украина
feldman@r5.dgtu.donetsk.ua, tanya@r5.dgtu.donetsk.ua

Вероятностная модель блочной памяти

Предложена вероятностная модель блочной памяти с одним и несколькими входными потоками, получены среднее число использованных банков и коэффициент загрузки, которое характеризует среднюю производительность модульной ОП.

Введение

Для реализации основной памяти используют объединение нескольких интегральных микросхем (ИМС). Совокупность микросхем называют *модулем памяти.* Один или несколько модулей образуют *банк памяти.*

В общем случае основная память вычислительного модуля практически всегда имеет блочную структуру, то есть содержит несколько банков.

При использовании блочной памяти, состоящей из B банков, адрес ячейки A *преобразуется* в пару (b, w), где b – номер банка, w – адрес ячейки внутри банка [1].

Схемы распределения разрядов адреса A между b и w:
- блочная (номер банка b определяет старшие разряды адреса);
- циклическая ($b = A$ mod B; $w = A$ div B);
- блочно-циклическая (комбинация двух предыдущих схем).

Типовая структура памяти, организованная в соответствии с блочной структурой, показана на рис. 1.

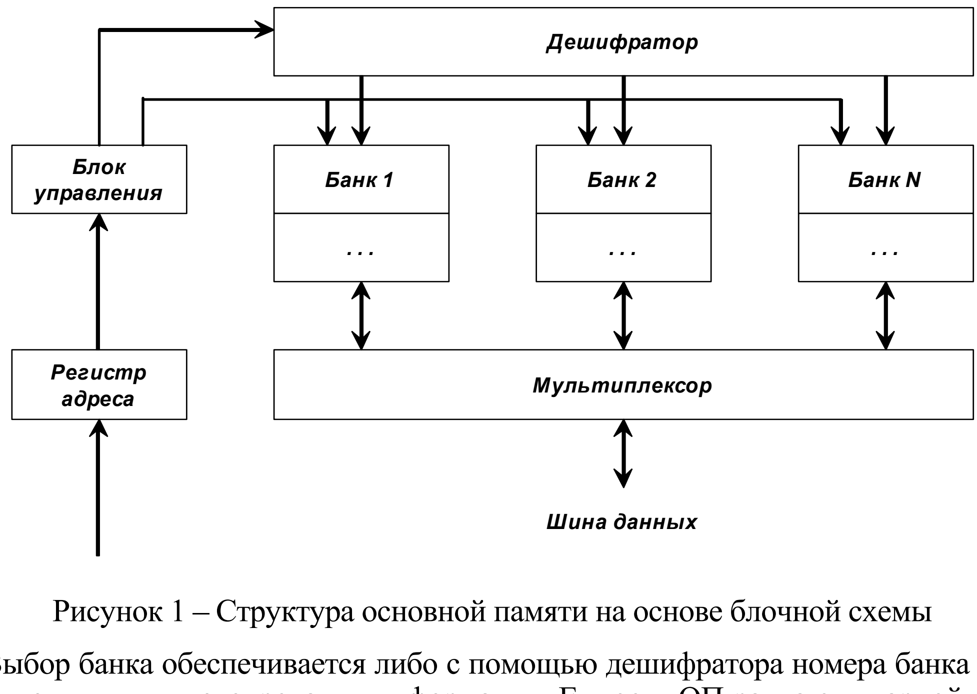


Рисунок 1 – Структура основной памяти на основе блочной схемы

Выбор банка обеспечивается либо с помощью дешифратора номера банка памяти, либо путем мультиплексирования информации. Емкость ОП равна суммарной емкости составляющих, а быстродействие – быстродействию отдельного банка [1].

Последовательный доступ в память обычно производится к ячейкам, имеющим смежные адреса. Можно достичь в N (N – количество банков) раз большей скорости обмена с памятью в целом, чем у отдельного ее банка, если обеспечить одновременное последовательное обращение к данным в каждом из банков.

В блочно-циклической схеме расслоения памяти каждый банк состоит из нескольких модулей, адресуемых по круговой схеме. Адреса между банками распределены по блочной схеме [1].

Целью работы является повышение эффективности работы блочной памяти на основе её аналитической модели с блочно-циклической схемой расслоения с одним модулем в каждом банке.

Вероятностная модель блочной памяти с одним входным потоком заявок

Основные предположения о функционировании блочной памяти, отражающие организацию работы ОП:

1. Все модули ОП работают синхронно так, что за один такт (цикл обращения) каждый модуль может обслужить не более одной заявки, а все модули не более N заявок одновременно.

2. В каждом такте работы число заявок во входном потоке не меньше числа банков системы, т.к. максимальная производительность всех существующих мультипроцессорных вычислительных систем достигается при вычислениях, производимых над регулярно организованными массивами информации. Число обращения к ОП за командами мало по сравнению с обращением за операндами. Потоки в рассматриваемой модели – это потоки обращения к ОП за считыванием или записью векторов, каждый из которых содержит элементов не менее, чем число банков ОП. Каждая заявка с равными вероятностями $1/N$ требует обслуживания каждым из банков ОП.

3. Все заявки, которые могут быть обслужены в одном такте работы ОП, поступают на обслуживание в соответствующие банки мгновенно, т.к. время пересылки одной заявки в несколько раз меньше цикла памяти.

4. Распределение заявок по банкам памяти происходит в порядке очередности этих заявок во входном потоке, из которого на обслуживание заявки поступают до тех пор, пока не появится первая заявка с запросом на обслуживание занятым банком. Это значит, что если какая-либо заявка не может быть обслужена в данном такте из-за занятости соответствующего банка, то она блокирует поступления к банкам ОП всех последующих заявок.

Интервал времени распределения и обслуживания заявок назовем тактом работы системы. Обозначим n число заявок входного потока, выбранных на обслуживание в одном такте, которое примем за состояние системы. Вероятность того, что первая заявка из входного потока поступит на обслуживание, равна 1, так как она застает свободными все N модулей. Вероятность p_1 того, что из входного потока на обслуживание поступит точно одна заявка, может быть вычислена как произведение вероятности того, что первая заявка поступит на обслуживание в один из N свободных банков, и вероятности того, что вторая заявка требует обслуживания тем же банком, т.е. $p_1 = 1/N$.

Вероятность того, что из входного потока на обслуживание поступит ровно две заявки, равна произведению вероятности поступления двух первых заявок на свободные модули и вероятности того, что третья заявка требует обслуживания одним из двух уже

занятых банков, т.е. $P_2 = ((N-1)/N) \times (2/N)$. Рассуждая аналогично, получим для вероятности p_n того, что из входного потока поступит на обслуживание ровно n заявок, формулу:

$$P_n = \frac{N-1}{N} \cdot \frac{N-2}{N} \cdots \frac{N-(n-1)}{N} \cdot \frac{n}{N} = \frac{N-(n-1)!}{(N-n)!N^n}, \quad n = \overline{1,N}. \tag{1}$$

Таким образом, на множестве состояний $\{O, N\}$ системы с заданным числом N обслуживающих устройств банков ОП определена цепь Маркова с переходными вероятностями p_n, $n \in \{O, N\}$, задаваемыми (1). Граф состояний цепи Маркова приведен на рис. 2.

Среднее число банков, занятых обслуживанием в течение одного такта, можно определить по формуле:

$$N_{cp} = \sum_{n=1}^{N} np_n = (N-1)! \times \sum_{n\text{-}1}^{N} \frac{n^2}{(N-n)!\,N^n}. \tag{2}$$

Значения среднего числа занятых банков $N_{\text{ср}}$ приведены в табл. 1. Коэффициент использования ОП, равный $R_{\text{ср}} = N_{\text{ср}}/N$, падает с ростом N и для количества банков больше восьми достаточно низкий.

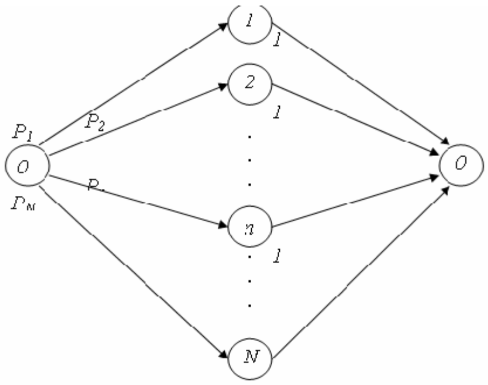


Рисунок 2 – Граф состояний цепи Маркова вероятностной модели блочной ОП

Таблица 1

Показатель	Число банков					
	1	2	4	8	16	32
$N_{\text{ср}}$	1	1,5	2,22	3,25	4,70	6,77
$R_{\text{ср}}$	1	0,75	0,58	0,40	0,29	0,21

Вероятностная модель блочной памяти с несколькими входными потоками заявок

Пусть в ОП, состоящую из N одинаковых банков памяти, поступают L независимых параллельных потоков обращения к банкам ОП от устройств обработки и периферийных процессоров мультипроцессорной вычислительной системы. К ограничениям для предыдущей модели добавятся следующие: между потоками установлен абсолютный приоритет на обслуживание в ОП, убывающий с ростом номера потока; каждая заявка любого потока с равными вероятностями $1/N$ требует обслуживания каждым из N банков ОП; распределение заявок каждого потока по банкам ОП происходит в порядке очередности этих заявок в данном потоке. Если какая-либо заявка l-го потока, $l = \overline{1,L}$, не может быть обслужена в данном такте из-за занятости соответствующего банка, то эта заявка блокирует поступления к банкам ОП и всех последующих заявок l-го потока. Затем происходит распределение на оставшиеся свободные модули ОП заявок из $(l+1)$-го потока и так далее.

Состояние системы – число занятых банков n_l, после окончания распределения заявок из входного потока, с l-м приоритетом. Вероятность перехода из состояния n_{l-1} в состояние n_l определяется:

$$P\{n_{l-1} \to n_l\} = \frac{(N-n_{l-1})!\,n_l}{(N-n_l)!\,N^{R_l+1}}, \tag{3}$$

где $R_l = n_l - n_{l-1}$ – число заявок, выбранных из l-го потока; $P\{n_{l-1} \to n_l\}$ – вероятность выбора на обслуживание равно n_l заявок l-го потока при условии, что из потоков более высокого, чем l-й, приоритет выбрано на обслуживание n_{l-1} заявок.

На рис. 3 приведена граф-схема марковской цепи для модели ОП с N банками и L входными потоками заявок.

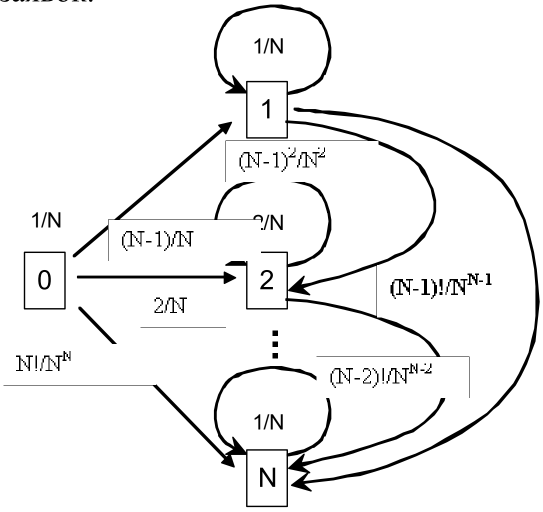


Рисунок 3 – Граф состояний цепи Маркова ОП с N банками и L входными потоками

В начальном состоянии цепи Маркова все модули памяти свободны, l-й шаг цепи Маркова соответствует распределению заявок L-го входного потока.

Вероятность перехода $P\{n_{l-1} \to n_l\}$ цепи Маркова определяется самими значениями $n_{l-1} = l$ и $n_l = j$.

$$p_{ij} = \begin{cases} \dfrac{(N-l)!\,j}{(N-j)!\,N^{j+l+1}}, & \text{для } l \le j,\ l = \overline{0,N} \\ 0 & \text{, для } l > j \end{cases} . \tag{4}$$

Этот граф состояния можно использовать для любого числа L входных потоков заявок. Матрица переходных вероятностей цепи Маркова, граф состояний которой изображен на рис. 3, имеет вид

$$P = \begin{bmatrix} 0 & P_{01} & P_{02} & \dots & P_{ON} \\ 0 & P_{11} & P_{12} & \dots & P_{1N} \\ 0 & 0 & P_{22} & \cdots & P_{2N} \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots & \vdots \\ 0 & 0 & 0 & \cdots & P_{NN} \end{bmatrix}.$$

Для каждого состояния марковской модели определены следующие стационарные вероятности:

$\pi_n(l)$ – вероятность застать на l-м шаге загруженные n банков памяти,

$\overline{\pi}(l) = (\pi_0(l), \pi_1(l), ..., \pi_N(l))$ – вектор вероятностей состояний на l-м шаге, $l = \overline{0,L}$. Поскольку начальное состояние известно, то известен и начальный вектор вероятностей состояний

$$\pi(0) = (1, 0, 0, \quad ..., 0).$$

Вектор вероятностей состояний на любом шаге $l+1$ может быть найден по рекуррентной формуле:

$$\overline{\pi}(l+1)=\overline{\pi}(l)P=\overline{\pi}(0)P^{l+1} \tag{5}$$

или в скалярной форме

$$\pi_n(l+1)=\sum_{l=1}^{N}\pi_l(l)p_{ln},\ n=\overline{1,N}. \tag{6}$$

Если определена вероятность $\pi_n(L)$ состояния n банков, занятых на L-м шаге, то можно определить и среднее число банков, занятых обслуживанием

$$N_{\text{ср}}(L)=\sum_{n=1}^{N}n\,\pi_n(L), \tag{7}$$

которое характеризует среднюю производительность блочной ОП с блочно-модульной схемой расслоения.

В табл. 2 приведены значения среднего числа занятых банков, из которых видно, что быстродействие памяти при одном и том же числе банков увеличивается вместе с ростом независимых потоков запросов. Однако увеличение производительности ОП, получаемое за счет разделения общего потока заявок на L независимых потоков, требует существенного усложнения управления работой мультипроцессорных вычислительных систем.

Таблица 2

Число банков	Число потоков							
	1	2	3	4	5	6	7	8
2	1.50	1.75	1.86	1.94	1.97	1.98	1.99	2.00
4	2.29	2.83	3.19	3.42	3.58	3.69	3.77	3.83

Выводы

Предложена вероятностная модель блочной памяти с блочно-модульной схемой расслоения с одним и несколькими входными потоками, которая позволяет определить среднее число занятых блоков и коэффициент использования, позволяющие оценить эффективность блочной памяти.

В дальнейшем эту модель можно использовать для анализа эффективности блочно-циклической оперативной памяти с несколькими модулями в банке.

Литература

1. Цилькер Б.Я. Организация ЭВМ и систем / Б.Я. Цилькер, С.А. Орлов. – М : ПИТЕР, 2004. – 668 с.
2. Столингс У. Структурная организация и архитектура компьютерных систем / Столингс У. – М. : Вильямс, 2002.– 893 с.

Л.П. Фельдман, Т.В. Михайлова

Ймовірнісна модель блочної пам’яті

Запропонована ймовірнісна модель блочної пам’яті з одним і декількома вхідними потоками, отримані середнє число використаних банків та коефіцієнт завантаження, яке характеризує середню продуктивність модульної ОП.

L. Feldman, T. Mikhaylova

Probabilistic Model of Block Memory

The probabilistic model of block memory with one and several input streams is suggested, the average number of used banks and factor of loading, which characterizes average capacity of module RAM are determined. The speed of memory under the same number of banks increases with the rise of independent flow of requests. However increasing of RAM capacity, which is got because of division of the general flow of the applications on L independent flows, requires the essential complication of management multiprocessor computing systems.

Статья поступила в редакцию 14.07.2010.